PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G06F 17/60

A1

(11) International Publication Number: WO 00/51045

(43) International Publication Date: 31 August 2000 (31.08.00)

(21) International Application Number: PCT/US99/28860

(22) International Filing Date: 6 December 1999 (06.12.99)

(30) Priority Data:
09/258,493 26 February 1999 (26.02.99) US

(71)(72) Applicant and Inventor: STEWART, Lisa, E. [US/US]; 1924 Walnut Street, Berkeley, CA 94704 (US).

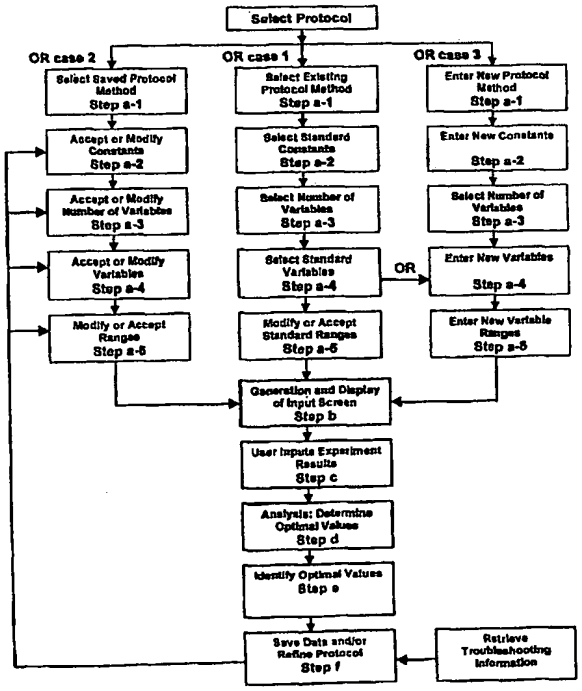
(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: EXPERT SYSTEM UTILIZING A KNOWLEDGE BASE AND DESIGN OF EXPERIMENT (DOE) TECHNIQUES

(57) Abstract

This invention provides the user with a design and optimization tool for the optimization of protocols. It integrates the advantages of a user interface with a simple set of steps, powerful statistical methods, flexible and specialized knowledge bases which can incorporate previously optimized protocol methods as well as standardized baseline data, measurement unit conversion tools, display of the experiment design analysis in a way which is easily understood, and an optional feedback method for the refinement of the protocol method.



BEST AVAILABLE COPY

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

TITLE

Expert system utilizing a knowledge base and design of experiment (DOE) techniques

BACKGROUND

FIELD OF THE INVENTION

This invention relates generally to expert systems and more particularly to an expert system for the optimization of protocols.

BACKGROUND OF THE INVENTION

The statistical methods of multivariable testing, also known as Design of Experiments (DOE), have been used in industrial process design for decades. However, it has not been embraced in the scientific community despite the significant advantages these techniques offer. One notable explanation for this is that DOE methods are perceived as formidably complex.

Scientists commonly design experiments using the traditional one-variable at a time approach. More specifically, all but one variable are held constant while the one under investigation is varied. The test variable is then fixed at some "good" value based on the results and another variable is modified.

This traditional method can be costly in both time and physical resources, particularly in cases where there is there is a wide variability in assay precision and linearity (a direct relationship) between variables. Additionally the traditional method does not evaluate the interactions among variables.

The statistical methods of DOE are very powerful techniques that can significantly enhance the

effectiveness of an experimental design. DOE methods can simultaneously consider interactions between many variables. DOE matrices can reduce the number of test configurations, reduce defects, improve experimental times, reduce expenses, improve the quality of experimental results, and greatly increase the odds of identifying hard-to-find solutions to diffucult quality problems. Thus it can be seen that DOE methods are cost effective in both time and physical resources. Further, quality experimental information translates into more reliable decisions and, ultimately, shorter times to product introductions.

Genichi Taguchi carried out significant research with DOE techniques in the late 1940's. His effort has been to make this powerful experimental technique more user friendly and apply it to improve the quality of manufactured products. Most of Taguchi's orthogonal arrays are easier-to-use rearrangements of earlier DOE designs. Interactions can be designed in and analyzed more easily, and the arrays can be modified for mixed-level designs with simple-to-follow steps.

Additionally, classical DOE does not specifically address quality. DOE using the Taguchi approach attempts to improve quality, which is defined as the consistency of performance. The prime motivation behind the Taguchi experiment design technique is to achieve reduced variation, also known as robust design. Robust designs, using ideas derived from Taguchi, allow the user to simultaneously study the controllable factors and reduce the effect of uncontrollable environmental variables. This technique, therefore, is focused on attaining the desired quality objectives in all steps.

Dr. Taguchi's standardized version of DOE, popularly known as the Taguchi method or Taguchi approach, was introduced in the USA in the early 1980's. Because of their simplicity and success in industrial process design, the Taguchi methods offer a cost-effective strategy involving interactions between wide ranging variable combinations. Today it is one of the most attractive quality building tools used by all types of engineers in the manufacturing industries.

The Taguchi philosophy of design of products and/or processes identifies three design stages: systems design, parameter design, and tolerance design. In the first stage, systems design, the designer draws upon his/her knowledge of the process in question to produce and initial design of a product or process. The use of experimentation may be irrelevant during this phase, but will become an essential element at the next stage, parameter design. The objective of parameter design is to choose suitable values for the parameters of the product or process. In the third stage, tolerance design, inexpensive components are replaced by better ones to achieve quality within the desired tolerance.

Current DOE tools assume that the user will have sufficient information to effectively define the first

two stages; systems design and parameter design. That is, that the user is familiar with the nuances of the particular plan of the scientific experiment or treatment (protocol method) being employed. They also assume that the user will be able to select appropriate constants, variables, and variable value ranges.

In addition, the user is expected to have a sophisticated knowledge of statistical design and analysis; many programs provide no guidance in the analysis of the results. Often, the user is presented with a lot of statistical output that requires substantial effort to translate results data into relevant answers.

In sum, the use of DOE tools can be complex, daunting, and can require a significant amount of time and effort to master.

These limitations make the use of current DOE design applications unattractive to those who could otherwise make productive use of these powerful statistical tools to significantly enhance the design, troubleshooting, and optimization of a laboratory protocol.

Other inventors have created several types of expert systems for protocol design, troubleshooting, and/or optimization employing DOE methods. However, none integrate a combination of a simple user interface, measurement unit conversion tools, specialized learning knowledge bases, a data structure for storing user tested protocol methods, a rule set which is used to process saved data and incorporate it into the knowledge bases, a hierarchy of parameter selection rules, robust experimental design and analysis tools, display the experiment design analysis in a way which is easily understood, and a feedback method for the refinement of the protocol method.

In addition many of these inventions are of such a sophisticated nature that their implementation is limited to hardware and software systems with specialized tools and are thus limited to a small group of users who have access to such facilities.

US Patent 4,472,770 (Li, September 18, 1984), US Patent 4,710,864 (Li December 1, 1987) and US PATENT 4,910,600 (Li March 20, 1990), "Self-optimizing method and machine", make use of statistical design matrix for automated experiment design and testing an object but assumes that this object is well defined, determines the test designs without human control or interaction, and does not integrate a knowledge base, nor does it have the ability to save results for future reference by others.

Patent JP7200662 (Hiroko, August 04, 1995), "Experiment Plan Production Support Systems Having Design Knowledge Base", requires that the product, the results of a completed process, has already been generated and that the relevant parameters of the initial process that produced the resulting

product are known. These results are required before the experiment plan can be generated. In addition, it does not provide for an optimizing feedback loop.

US Patent 5,107,499 (Lirov, et al. April 21, 1992), "Arrangement for automated troubleshooting using selective advice and a learning knowledge base", interactively communicates between a user and utilizes a learning knowledge base but it does so in a complex fashion and does not incorporate DOE design methods.

US Patent 5,253,331 (Lorenzen, et al. October 12, 1993), "Expert system for statistical design of experiments", defines a method for interacting with a user to specify and experimental design. However, it does not utilize a knowledge base, nor provide a feedback method after the experiment has been completed, and involves complex interactions between multiple layers of programming language tools and is thus is restrictive in the type of computer hardware and software platforms on which it can be developed.

REFERENCES

- "Statistics in Research and Development", Second Edition, Ronald Caulcutt, Chapman & Hall, 1991
- Cobb, B. D. and J. M. Clarkson (1994), "A Simple Procedure for Optimizing the Polymerase Chain Reaction (PCR) Using Modified Taguchi Methods," Nucleic Acids Research, Vol. 22, No. 18, pp. 3801-3805.
- Briones, P., "Experimental Design: A useful Tool for PCR Optimization", BioTechniques, 21:134-140
- "Why Don't More Researchers Use Design of Experiments?", R&D Magazine, January 1995, pp. 31
- "DOE Makes Research Pay Off", R&D Magazine, April 1997, pp. 43

SUMMARY

In accordance with the present invention, an expert system is provided which is comprised of a simple user interface, flexible and specialized learning knowledge bases, data structures and process for storing user tested protocol methods, hierarchy of parameter selection rules, measurement unit conversion tools, robust experimental design and analysis tools, display of the experiment design analysis in a way which is easily understood, and an optional feedback method for the refinement of the protocol method.

OBJECTS AND ADVANTAGES

Accordingly, several objects and advantages of the present invention are that it:

- Has a simple design for ease of use; it requires the user to complete only a few straightforward selection steps.
- Allows the user to store tested protocol design results in a specialized knowledge base which can later be accessed by others.
- Allows the user to easily identify standard and previously optimized protocol method constants, variables, and variable ranges by the use of a learning knowledge base. Thus an inexperienced user can efficiently make use of the experience of others which is stored in the knowledge base.
- Requires little previous experience with statistics to utilize powerful statistical optimization tools. Thus it creates an opportunity for users who would normally not take advantages of powerful DOE tools because of a lack of the time and/or skills to do so.
- Has the flexibility of allowing the user to input and test novel protocol methods, constants, variables, and variable ranges and to add them to the shared knowledge base.
- Allows dependent variables with linear (direct) and non-linear relationships to other variables and constants can be quickly and easily identified and the appropriate values determined from the knowledge base.
- Allows users to easily refine a protocol method based on previous findings using a feedback selection and repeating the experiment with the best values identified in the analysis step.
- Can more easily implemented on a wide range of current and emerging computer hardware and software systems, unlike other expert systems, because of its simplicity of design.

Other objects and advantages of this invention are that it:

- Saves time and labor because it minimizes the amount of effort a user must apply to collect preliminary protocol data, design an experiment, and analyze the results.
- Saves time and labor because it enables the user to draw upon work previously completed by others; it reduces duplicated efforts.
- Saves time and labor because it minimizes the number of trials necessary to troubleshoot and/or optimize a protocol method.
- Satisfies a need for a fast and efficient method of optimizing new and evolving laboratory protocol methods for such quickly growing and competitive industries as biotech, healthcare, and pharmaceuticals.
- Takes into account the wide variation of measurement unit standards in many fields of science by incorporating convenient measurement unit conversion tools for experimental constant and variable range units to standard and non-standard values.
- Has the flexibility to incorporate knowledge bases of varied structure for efficient management of a broad range of protocol data.

Further objects and advantages of this invention will become apparent from a consideration of the drawings and the ensuing description

DRAWING FIGURES

FIG. 1 is a generalized flow chart of the present invention.
FIG. 2 is a generalized series of selection/input screen elements for steps a
FIG. 3 is a generalized series of selection/input screen elements for steps b through f.
FIG. 4 is a generalized flow chart of the relationships among the knowledge base data.

REFERENCE NUMERALS IN DRAWINGS

11 Step a-1: selection/input of a protocol and protocol method.
12 Step a-2: selection/input of constants.
13 Step a-3: selection of DOE matrix design and/or number of variables and variable value levels to be tested.
14 Step a-4: selection/input of variables.
15 Step a-5: selection/input of variable ranges.

16 Step b: Generation and display of experiments and results input screen elements.
17 Step e: Identification of optimal values, troubleshooting, data storage, and step refinement options.

50 Structures: Protocols data
52 Structures: Methods data
54 Structures: Constants data
56 Structures: DOE matrix design data
58 Structures: Variables data
60 Structures: Variable range data
62 Process: Results data input and Analysis of test results
64 Structures: Saved user tested methods data
66 Process: Analysis of collected user methods data
68 Structures: Troubleshooting data
70 Structures: Parameter relationships where parameters include all factors in a protocol

CONCLUSION, RAMIFICATIONS, AND SCOPE OF INVENTION

This invention provides the user with a unique tool which has the advantages of a simple user

interface, flexible and specialized learning knowledge bases, data structure and process for storing user tested protocol methods, hierarchy of parameter selection rules, measurement unit conversion tools, robust experimental design and analysis tools, display of the experiment design analysis in a way which is easily understood, and an optional feedback method for the refinement of the protocol method.

There are many novel and emerging protocol methods whose development could be significantly enhanced by the current invention. An example is the polymerase chain reaction (PCR). An important emerging biological protocol, PCR is a technique that is used to make many copies of a specific region of DNA. This is done to produce enough copies so that the DNA can be adequately tested. It has a rapidly growing range of uses in such areas as assessment of therapy effectiveness, the detection of infectious microorganisms, genome mapping, quality testing, basic molecular biology research, and many other significant applications. PCR has a wide variability in assay (analysis of one or more components) precision and linearity between the various methods and standards; as with any new technique, there are a wide range of standards used by scientists and described in the literature. In addition, because PCR is a relatively new protocol, a large number of people using it have a limited knowledge of PCR methods. The current invention addresses these issues.

More generally this application can be applied to such areas as medicine. For example, in the optimization of treatments which involve complex drug and therapeutic combinations and interactions. In business and marketing, professionals could use this tool to improve their products and services by determining what feature combinations are most needed and desired by consumers. It could be used to define effective rules and policies by helping to define what combination of factors influence a particular outcome in social environments.

We are in an era where data and information are accumulating at an accelerating rate. The explosive power of the Internet, which this invention is designed to utilize, is a notable example. To effectively and efficiently take advantage of this growing base of information, there is a need for tools such as this invention. These tools simplify the analysis of complex combinations of information and determine the significant factors so that effective and reliable decisions can be made.

The potential for an invention such as this is enormous. Because of its relatively simple design, relevant information can be identified and utilized with substantially more ease and convenience than before. Because of the powerful statistical tools that it incorporates, information can be uncovered that was previously unknown about the interactions between variable element combinations. People who would otherwise be limited in the scope of their contributions to the advancement of research and

analysis of trends can participate more fully in the discovery of this knowledge.

While the above description contains many specifics these should not be construed as limitations on the scope of the invention, but rather as an exemplification of one preferred embodiment thereof. Many other variations are possible.

Accordingly, the scope of the invention should be determined not by the embodiments illustrated, but by the appended claims and their legal equivalents.

DESCRIPTION Figs 1-5

Description of Fig 1
A generalized flow chart of the present invention is illustrated in **Fig 1**. The user begins by selecting one of three options to define his/her experiment design parameters:

case 1: with the guidance of the knowledge base and help contents
case 2: by modifying a previously designed, tested, and retrieved protocol method
case 3: by inputting entirely new protocol method data.

A typical embodiment of the method of the present invention begins at **step a-1**. In **case 1** the user selects from a list of currently defined protocol methods available in the database; in **case 2** the user makes no changes; in **case 3** the user inputs the name of the new protocol method to be tested.

For **step a-2** the user identifies the constants that are specific to the method selected in **step a-1**. In **case 1** constant selections are retrieved from the knowledge base; in **case 2** the user either accepts or modifies existing constants; in **case 3** the user inputs the appropriate constants.

Next in **step a-3**, the user selects the number of variables to be tested. This selection determines the design matrix that will be used in generating the experiment design. In the preferred embodiment the robust design matrixes of Taguchi are used. In **case 1** the user selects from the knowledge base options; in **case 2** the user either accepts or modifies the existing number of variables; in **case 3** the user selects from the DOE method options.

In **step a-4** the user identifies the variables specific to the method identified in **step a-1**, to the constants identified in **step a-2**, and the number of variables which have been determined in **step a-3**. In **case 1** variable selections are retrieved from the knowledge base; in **case 2** the user either accepts

or modifies existing variables; in **case 3** the user inputs the appropriate variables.

In **step a-5** the user identifies variable range values specific to the method identified in **step a-1**, to the constants identified in **step a-2**, the number of which have been determined in **step a-3**, and specific to the variables selected in **step a-4**. In **case 1** variable range selections are retrieved from the knowledge base. These selections can either be standard protocol method values or determined by functions based on the relationships between the variables and constants identified in the preceding selections; in **case 2** the user either accepts or modifies existing variable ranges; in **case 3** the user inputs the appropriate variable ranges.

After the parameter selection just described, all of the three cases noted above now follow the same series of steps.

In accordance with the input of the data identified in **steps a-1** through **a-5**, an experiment design is generated in **step b**.

The user now inputs the results of the experiment design in **step c**. Alternately, the results data can be retrieved from an automated system which controls the experiment tests and/or measures the test results.

The next action is the analysis of the data input in **step c**. For the Taguchi method in the preferred embodiment, the function used in **step d** is the signal to noise ratio or SNT:

$$SNT = -10 \log[(SD/mean)^2]$$

Where SD is the standard deviation. The mean is a middle point between extremes; that is, the expected value.

In **step e**, the optimal combination of variable values for this experimental design is identified and displayed for the user. In the preferred embodiment, a display of experimental variables and their respective range values are displayed with the best combinations highlighted so that the user can easily interpret the results.

Finally, in **step f** the user can save the experiment and the test results to the knowledge base. This data is saved so that it can later be retrieved for use by others and/or additional refinement. In addition, the protocol method parameters and related test results can be processed and incorporated into the standardized knowledge base as appropriate.

The user may now return to any of the previous steps to refine his/her experiment protocol based on the information gathered in the preceding process. In the preferred embodiment if the user returns to **step a-5**, the selection of variable ranges, the best values identified in **step e** are now the proffered mid-range values. The higher and lower values can either be determined by a set of standard default ranges or by formulas which define known linearities between interrelated constants and variables. These ranges and formulas will be unique to the protocol method, constants, and/or variables involved.

Because of the wide range of protocols, each with its own unique components and component relationships, there are various possibilities with regards to the specific interface designs for **step a**, the definition of experiment parameters.

Knowledge base schemas will vary for this reason as well. More specifically, the relational database designs, which are composed of individual structures containing protocol methods, constants, variables, variable ranges, parameter relationships, and troubleshooting data will depend on the protocols being addressed. A generalized database schema is illustrated in **Fig 4**.

Description of Fig 2

A generalized illustration of selection/input and display screen elements are presented in **Figs 2** and **3**. In the preferred implementation, the knowledge base data is stored on servers which can be accessed over a LAN (Local Area Network), WAN (Wide Area Network), or the Internet. In this way many people can easily and conveniently share knowledge base data. In all of the illustrations, the boxes with a "?" indicate a help button. The help buttons allow the user access to relevant parameter detail information in the knowledge base.

On the left-hand side of each of the **step a** selection/input screen elements in **Fig 2** there is a selection pop-down box. These are the selections determined by the knowledge base and applied in **case 1** and **case 2** discussed and identified in **Fig 1**. On the right hand side of the selection/input screens there are boxes which identify input boxes. These are for **case 3** in which the user inputs new information which is not present in the knowledge base.

The following description will use PCR as a sample protocol to help describe how the illustrated input screens can relate to one another. The following will be **case 1**, a new experiment design which is made with selections from the knowledge base.

In **step a-1** (11) the user selects a protocol, PCR, and a method. The method selections are

determined by the protocol selection. Because there may be many methods, each with it's own advantages and disadvantages, the user can optionally review the necessary method information retrieved from the knowledge base by clicking on the help button. In this example the PCR protocol method selected will be *in situ* PCR, a knowledge base selection from the pop-down menu.

The method selected will determine which set of constants the user selects from in **step a-2** (12). In this case the constants would be polymerases, enzymes which catalyze the formation of DNA, and primer pairs, short sequences of DNA which bind to the ends of the sequence to be copied. The selection of these constants are guided by information provided to the user from the knowledge base by clicking on the help button.

These selected constants will later influence the variables and variable ranges suggested to the user.

The number of variables to be tested are determined next in **step a-3** (13). In this example, the Taguchi method is selected and the user identifies that he/she wants to test four variables at three levels (min, mid, and max variable value levels). The Taguchi method uses specific matrix sizes which will determine the number of tests to be run. For this reason, the number of variables identified by the user determines the experimental design matrix used. Alternately, the matrix design selected will determine the number of variables and variable values to be tested. The user is guided in this selection by information which can be retrieved from the knowledge base by clicking on the help button.

Based on the PCR method and constants selected, the user is presented with a set of variables to choose from in **step a-4** (14). The number of variables the user can select from are those he/she identified in **step a-3** (13). In this case, the user will select four variables. The user is guided in this selection by information which can be retrieved from the knowledge base by clicking on the help button. Alternately the user can enter the name/id of other variables he/she would like to test which are not currently in the knowledge base.

If in the aforementioned steps the user has selected values that exist in the knowledge base, variable value levels will be suggested to be accepted or modified in **step a-5** (15). In this example, the suggested variable value levels for each of the four selected variables will appear in the boxes labeled "Level 1", "Level 2", and "Level 3".

For example, if the protocol method chosen is in situ PCR, the variable Mg++ has a standard range of 1.5 - 5.0 mM. However, if the polymerase constant chosen is the polymerase UlTma, the suggested Mg++ range will be limited to 1.5 - 2.0 mM. Thus the three boxes would display "1.5 mM", "1.75 mM",

and "2.0 mM" respectively.

Similarly, the standard melting temperature range is 68°C - 75°C. However, the suggested temperature ranges will be determined with a formula which uses the characteristics of the primer pair constants: 62.3°C + 0.41°C (%G-C) - 500/(primer length).

In either case, the user can utilize a unit conversion tool so that the units are most convenient for the user. An example would be if the user prefers to use mg/L instead of mM. By changing the pop-down selection from the default measurement units of mM to mg/L, the values in the three boxes would be converted to the appropriate values for the selected unit.

The suggested values in these boxes can also be modified by the user as needed.

If the variable is not currently identified in the knowledge base, the user will enter the ranges known by him/her to be reasonable in this step.

Description of Fig 3
Now that the user has made the necessary parameter selections in **steps a**, the experiment design and the input screen elements for experiment results in **step b** (16) are generated. In **step c** the user inputs the results of the experiments described in the table. In this example the set of nine experiments displayed are those of a Taguchi matrix for four variables and three range levels.

After the analysis of these results are completed in **step d**, the optimal combination of experimental variable values are identified in a clear and simple manner to the user in **step e** (17). In this example, the best variable levels are identified by the bold boxes.

The user can now get troubleshooting information by clicking on the help button. If, for example, the user has low specificity in his/her PCR reaction, he/she can identify possible causes and solutions from the information retrieved from the knowledge base.

In **step f** the user can click on the "SAVE" button to save the data from the completed process to the knowledge base for future reference. This data can also be processed along with other saved protocol method test results to determine linear and non-linear relationships between parameters and incorporated into the knowledge base data used for the initial design of protocols.

If further refinement of the protocol method is desired, the user can return to any of the previous steps. If the user returns to **step a-5** (15) directly from **step f** (17), the optimal values determined in the just

completed series of steps will be the new mid-range values. In this example the best variable value levels indicated with bold boarders in the analysis screen (17) will appear in the boxes labeled "Level 2". New higher and lower variable range values can be estimated from previously defined rules from the knowledge base and accepted or modified by the user. These rules will be unique to the protocol method, constants, and/or variables involved.

Description of Fig 4

Introduction: This figure is a generalized knowledge base schema to help illustrate the preferred embodiment discussed. The data structures can include, for example, database tables or data arrays. Any of the data types identified by the boxes in the figure can consist of a number of related data structures. In this way, data of the same type, such as constant data (54), with unique characteristics can be stored and retrieved more efficiently.

The parameter relationship structures (70) identify the known optimal range values for data combinations. It could be static maximum and minimum values or it could be a formula based on a data element combination such as a selected method and a variable. For example if the data pair is the method *in situ* PCR and the variable Mg++ the known optimal variable range for Mg++ would be identified as the static values 1.5 mM to 5.0 mM.

Similarly for the case where the data pair is the constant primers and the variable temperature. The relationship will be defined by a formula based on the characteristics of the selected primer pair: 62.3°C + 0.41°C (%G-C) - 500/(primer length) will be the suggested melting temperature. Note that in some cases, there can be relationships between more than two parameters, such as two variables and a constant, which this set of parameter relationship data can identify. This is illustrated in the case of two variables Mg++ and dNTP and the constant polymerase UlTma. To determine the best variable range for dNTP the relationship information indicates that Mg++ ions should exceed that of total dNTP concentration by 0.5 mM - 3.0 mM. As identified above, the Mg++ level for UlTma is best between 1.5 - 2.0 mM. Thus, the suggested range of dNTP, 2.0 – 5 mM, is determined by considering its combination with another variable and a constant.

Schema Description: The process begins by the selection from the protocols data (50) followed by a selection from the protocol methods data (52). In the example this would be the protocol PCR and the method *in situ*.

The possible and/or required constants used in this method are identified from the constants data (54). In the example, the constant types identified from the data structure would be polymerases and primer sets.

The constant elements of polymerases and primer sets identified are also chosen from this set (54); UlTma and a primer set AB. As noted above, there can be more than one data structure in a data type set. In this example the polymerase and primer set selections and related information are contained in separate but similarly defined data structures to facilitate efficient data storage and retrieval.

For the purposes of illustrating the preferred embodiment, the DOE matrix selection takes place next. This selection determines how many variables are selected in the next step. Alternately, the selection of the number of variables and range levels the user wishes to test will determine which DOE matrix is used for the design and analysis of the experiment.

The appropriate number and type of variable data (58) are now selected. These selections are determined by the previously selected parameters and the parameter relationship sets (70). For example, the variables presented for selection will be dependent on the method chosen and possibly the constant selections. In addition, the number of variables that can be selected are limited to the number identified in the matrix selection set.

For consistency with FIG 1, variable range data (60) is represented separately from the variable data structure (58) in this schema. Variable ranges are determined by referring to the parameter relationship data (70) that identify relationships between parameters such as the selected method, constants, and other variables.

After the experiment test results have been entered, it is analyzed (62). Here, if required, the troubleshooting data structure (68) is referred to as appropriate. The completed method experiment design, results, and analysis can now be stored in the saved method data structure (64).

The saved method data structure (64) can be in any number of formats, depending on the size of the data set and software/hardware platform being used. In the preferred embodiment it is stored in a flat file, such as a browser cookie. For larger and more complex protocol methods it may have any number of embodiments.

This saved method data (64) can be added to the knowledge base after being processed though the method data analysis (66). This analysis consists of a set of rules that determine the appropriate storage in any of the knowledge base data structures mentioned above.

CLAIMS

1. A method of operating a computer to enable said computer to execute a design, analysis, and optimization of a protocol, said method comprising the steps of:

a) inputting information to the computer, the information including data from the group consisting of data obtained from user input, data selected from a knowledge base comprising a plurality of protocol data, and experimental designs selected from the group consisting of a plurality of DOE methods.

b) generating and displaying an experiment design in accordance with the parameters selected in step a

c) inputting information to the computer, the information including the results of the tests performed in accordance with the experimental design generated in step b

d) evaluating the experiment results by utilizing the information input in step c and the DOE analysis method identified in step a

e) communicating the values computed in step d

whereby a user can simply, easily, and efficiently design, optimize, and/or troubleshoot a protocol by use of information from the group including data obtained from user input, a knowledge base comprising a plurality of protocol data, and experimental designs selected from the group consisting of a plurality of DOE methods.

2. The method of claim 1 wherein step e is followed by the provision of a feedback option to return to any of the previous steps for experiment refinement.

3. The method of claim 1 wherein said knowledge base is selected from the group consisting of scientific protocol information.

4. The method of claim 3 wherein said knowledge base consists of scientific protocols in the biological sciences.

5. The method of claim 4 wherein said knowledge base consists of scientific protocols relating to genomics.

6. The method of claim 1 wherein said experiment design is selected from the group consisting of robust design methods.

7. The method of claim 1 wherein said protocol data is selected from the group consisting of parameters comprised of characteristics, value ranges, and relationships of data from the group

including protocols, protocol methods, constants, and variables.

8. The method of claim 1 wherein any parameter selection is accompanied by a resource for communicating information selected from the group consisting of protocol data.

9. The method of claim 1 wherein any parameter selection is accompanied by a resource for converting said protocol data measurement units to measurement units selected from the group consisting of all measurement units.

10. The method of claim 1 wherein said knowledge base includes data processed by a rule set applied to data selected from the group consisting of completed protocol designs, results, and results analysis.

11. The method of claim 1 wherein data selected from the group consisting of data input by the user, said knowledge base data, rule set data, and analysis data can be output to a device selected from the group consisting of static presentation processors, storage devices, and display devices.

12. The method of claim 1 wherein the information communicated is communicated to the user with an object selected from the group consisting of visual display elements and audio elements.

13. The method of claim 1 wherein the knowledge base is distributed among more than one computer.

14. The method of claim 1 wherein said inputting of information to said computer is controlled by a system selected from the group consisting of automated systems.

15. A machine for the design, analysis, and optimization of a laboratory protocol, said machine comprising:

means for inputting information to the computer, the information including data from the group consisting of data obtained from user input, data selected from a knowledge base comprising a plurality of protocol data, and experimental designs selected from the group consisting of a plurality of DOE methods.

means for generating and displaying an experiment design in accordance with said DOE design and said parameters

means for inputting result data to the computer, the data including the results of the tests performed in accordance with the experimental design

means for evaluating said results data

means for communicating results analysis data, the results analysis data including the evaluation of the results of the tests performed in accordance with the experimental design

means for saving the completed test design and results for retrieval, evaluation, and incorporation into said knowledge base.

whereby a user can simply, easily, and efficiently design, optimize, and/or troubleshoot a laboratory protocol by means of information from the group including data obtained from user input, a knowledge base comprising a plurality of protocol data, and experimental designs selected from the group consisting of a plurality of DOE methods.

16. The machine of claim 15 which includes means for feedback of best results to any of the previous steps for protocol design refinement.
17. The machine of claim 15 wherein said laboratory protocol consists of scientific protocols in the biological sciences.
18. The protocol of claim 17 wherein said laboratory protocol consists of biological protocols relating to genomics.
19. The machine of claim 15 wherein said experiment design is selected from the group consisting of robust design methods.
20. The machine of claim 15 wherein said protocol data is selected from the group consisting of characteristics, ranges, and relationships of data from the group including protocols, protocol methods, constants, and variables.
21. The machine of claim 15 wherein any parameter selection is accompanied by a resource for communicating information selected from the group consisting of protocol data.
22. The machine of claim 15 wherein any parameter selection is accompanied by a resource for converting said protocol data value ranges to value ranges selected from the group consisting of all measurement units.
23. The machine of claim 15 wherein said knowledge base includes data collected and processed by a rule set applied to data selected from the group consisting of user input, completed protocol designs, results, and results analysis.
24. The machine of claim 15 wherein data selected from the group consisting of data input by the user, said knowledge base data, rule set data, and analysis data can be output to a device selected from the group consisting of static presentation processors, storage devices, and display devices.
25. The machine of claim 15 wherein the information communicated is communicated with an object selected from the group consisting of visual display systems, audio systems, and network systems.
26. The machine of claim 15 wherein the knowledge base is distributed among more than one computer.
27. The machine of claim 15 wherein said inputting of information to said computer is controlled by a system selected from the group consisting of automated systems.

28. A machine for the design, analysis, and optimization of a protocol, said machine comprising:

means for selecting a protocol said protocol selected from the group consisting of data obtained from user input and data selected from a knowledge base comprising a plurality of protocol data

means for selecting a method said method selected from the group consisting of data obtained from user input and data selected from a knowledge base comprising a plurality of said protocol method data

means for selecting constants said constants selected from the group consisting of data obtained from user input and data selected from a knowledge base comprising a plurality of said protocol method constants data

means for selecting variables said variables selected from the group consisting of data obtained from user input and data selected from a knowledge base comprising a plurality of said protocol method variable data

means for selecting variable value ranges said variable value ranges selected from the group consisting of data obtained from user input and data selected from a knowledge base comprising a plurality of said protocol method variable value range data

means for selecting experiment designs said experiment designs selected from the group consisting of a plurality of DOE methods.

means for generating and displaying an experiment design in accordance with said DOE design and said protocol data

means for inputting result data to the computer, the data including the results of the tests performed in accordance with said experimental design

means for evaluating said results data

means for communicating results analysis data said results analysis data including the evaluation of the results of the tests performed in accordance with the experimental design

means for saving the completed test design, results, and analysis for retrieval, evaluation, and incorporation into said knowledge base.

whereby a user can simply, easily, and efficiently design, optimize, and/or troubleshoot a laboratory

protocol by means of information from the group including data obtained from user input, a knowledge base comprising a plurality of protocol data, and experimental designs selected from the group consisting of a plurality of DOE methods.

29. The machine of claim 28 wherein the incorporation of a completed test design and results data are processed by a rule set the rule set selected from the group comprising data processing.

30. The machine of claim 28 wherein said method, constants, variables, variable ranges, and rule sets are selected from the group consisting of knowledge base data which is determined by a rule set selected from the group comprising data processing.

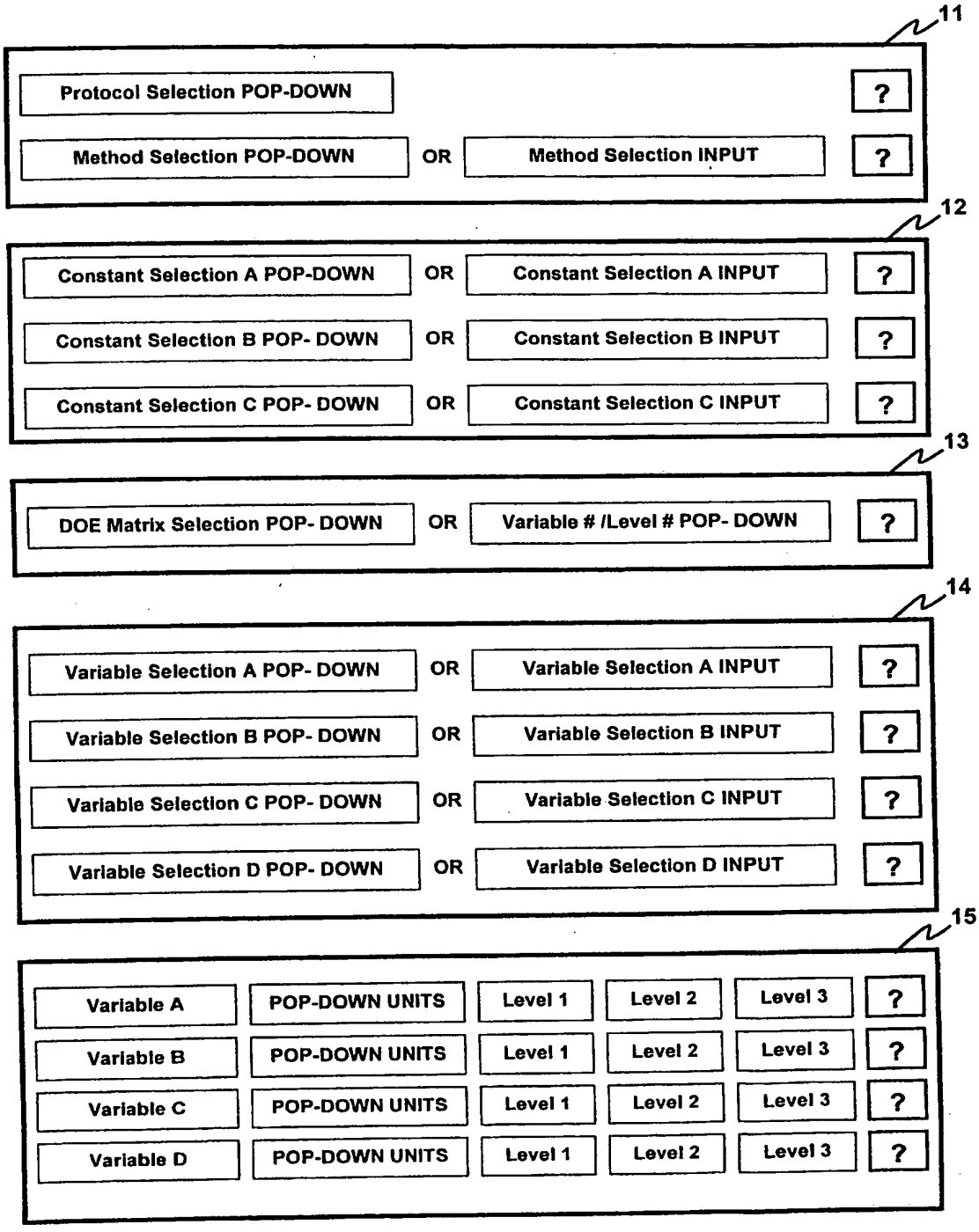

11
Protocol Selection POP-DOWN
?
Method Selection POP-DOWN
OR
Method Selection INPUT
?
12
Constant Selection A POP-DOWN
OR
Constant Selection A INPUT
?
Constant Selection B POP- DOWN
OR
Constant Selection B INPUT
?
Constant Selection C POP- DOWN
OR
Constant Selection C INPUT
?
13
DOE Matrix Selection POP- DOWN
OR
Variable # /Level # POP- DOWN
?
14
Variable Selection A POP- DOWN
OR
Variable Selection A INPUT
?
Variable Selection B POP- DOWN
OR
Variable Selection B INPUT
?
Variable Selection C POP- DOWN
OR
Variable Selection C INPUT
?
Variable Selection D POP- DOWN
OR
Variable Selection D INPUT
?
15
Variable A
POP-DOWN UNITS
Level 1
Level 2
Level 3
?
Variable B
POP-DOWN UNITS
Level 1
Level 2
Level 3
?
Variable C
POP-DOWN UNITS
Level 1
Level 2
Level 3
?
Variable D
POP-DOWN UNITS
Level 1
Level 2
Level 3
?

FIG 2

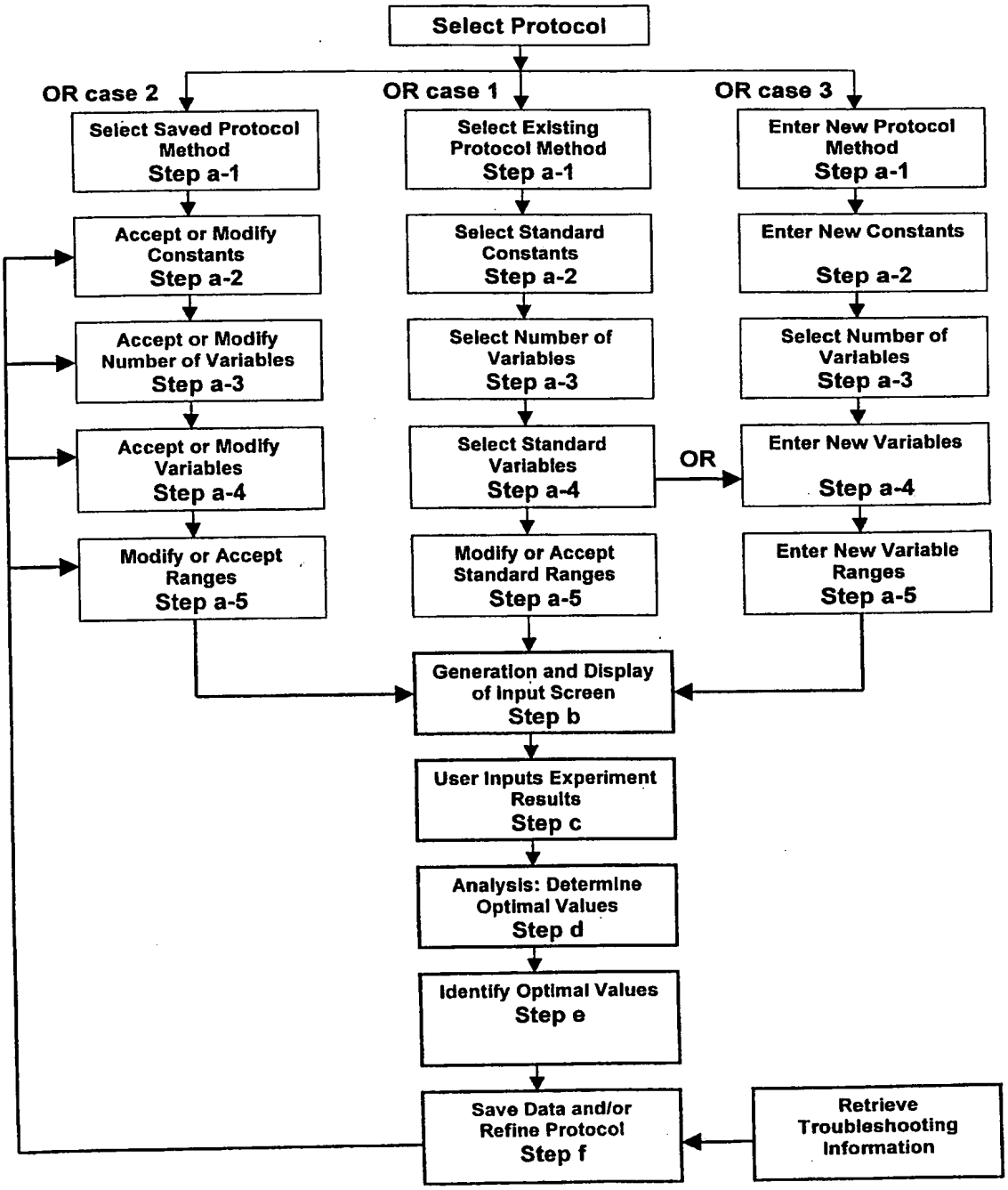

Select Protocol
OR case 2
OR case 1
OR case 3
Select Saved Protocol Method Step a-1
Select Existing Protocol Method Step a-1
Enter New Protocol Method Step a-1
Accept or Modify Constants Step a-2
Select Standard Constants Step a-2
Enter New Constants Step a-2
Accept or Modify Number of Variables Step a-3
Select Number of Variables Step a-3
Select Number of Variables Step a-3
Accept or Modify Variables Step a-4
Select Standard Variables Step a-4
OR
Enter New Variables Step a-4
Modify or Accept Ranges Step a-5
Modify or Accept Standard Ranges Step a-5
Enter New Variable Ranges Step a-5
Generation and Display of Input Screen Step b
User Inputs Experiment Results Step c
Analysis: Determine Optimal Values Step d
Identify Optimal Values Step e
Save Data and/or Refine Protocol Step f
Retrieve Troubleshooting Information

FIG 1

16

	Variable A	Variable B	Variable C	Variable D	
Experiment 1	Level 1	Level 1	Level 1	Level 1	Result 1
Experiment 2	Level 1	Level 2	Level 2	Level 2	Result 2
Experiment 3	Level 1	Level 3	Level 3	Level 3	Result 3
Experiment 4	Level 2	Level 1	Level 2	Level 3	Result 4
Experiment 5	Level 2	Level 2	Level 3	Level 1	Result 5
Experiment 6	Level 2	Level 3	Level 1	Level 2	Result 6
Experiment 7	Level 3	Level 1	Level 3	Level 2	Result 7
Experiment 8	Level 3	Level 2	Level 1	Level 3	Result 8
Experiment 9	Level 3	Level 3	Level 2	Level 1	Result 9

17

Variable A	Level 1	Level 2	Level 3	SAVE
Variable B	Level 1	Level 2	Level 3	
Variable C	Level 1	Level 2	Level 3	
Variable D	Level 1	Level 2	Level 3	?
Refine:	Step 2	Step 3	Step 4	Step 5

FIG 3

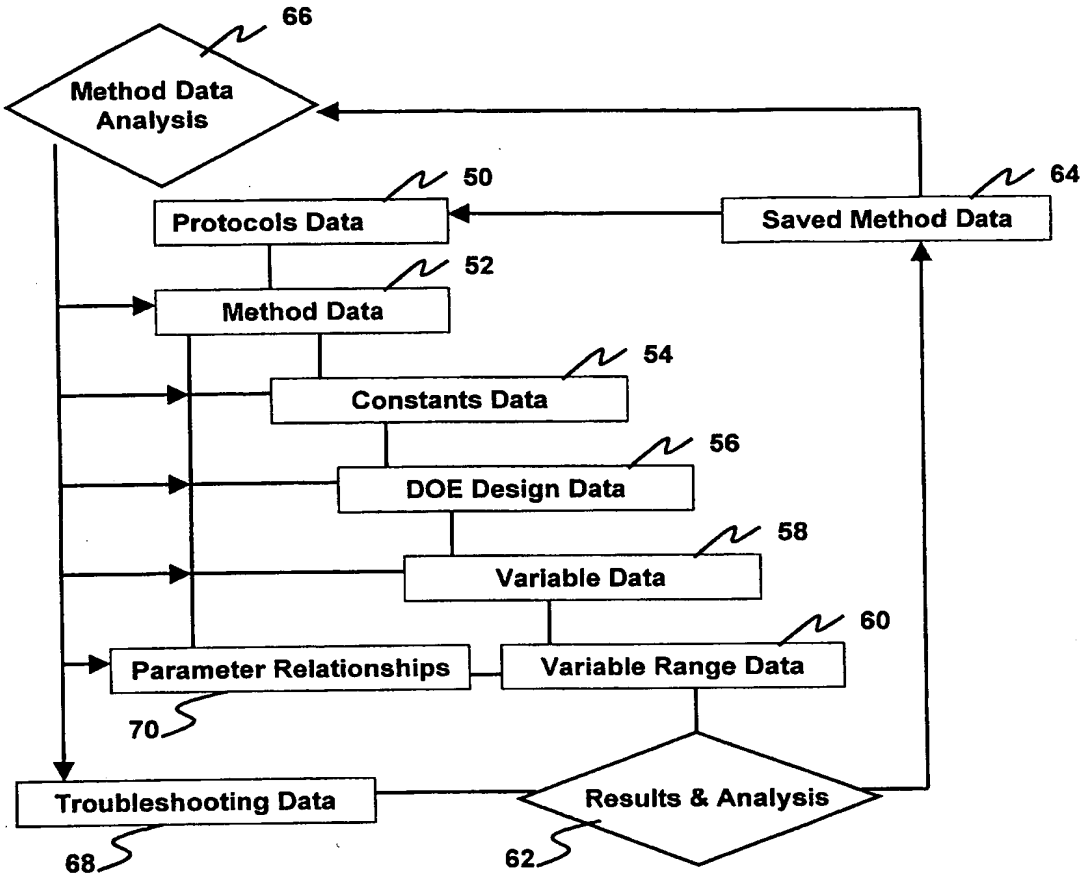

66
Method Data Analysis
50
Protocols Data
64
Saved Method Data
52
Method Data
54
Constants Data
56
DOE Design Data
58
Variable Data
60
Parameter Relationships
Variable Range Data
70
Troubleshooting Data
Results & Analysis
68
62

Fig 4

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/28860

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/60

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PINTO NABIL B, BONNELL RONALD D: "IPAD: An Intelligent Parameter Design Tool" PROC. 3RD INT. CONF. ON INDUSTRIAL AND ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE AND EXPERT SYSTEMS, 15 - 18 July 1990, pages 375-384, XP002136604 section 2.X: IPAD section 3.X: TAG -/--	1-30

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 April 2000	22/05/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Schmidt, A

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	MCEWAN W: "Taguchi methods and expert systems in fabrication design" INTERNATIONAL JOURNAL OF PRESSURE VESSELS AND PIPING, vol. 53, no. 1, 1993, pages 47-61, XP002136605 abstract section: Taguchi Methodology for Fabrication Designs section: Expert Systems in NDT	1-30
A	MCEWAN W ET AL.: "Parameter Design and Expert Systems in NDT" BRITISH JOURNAL OF NON-DESTRUCTIVE TESTING, vol. 33, no. 3, March 1991 (1991-03), pages 115-120, XP002136606 the whole document	1-30